An Introduction to Process Variability Simulation

Celso Bation Co
Research and Development Office
Technological Institute of the Philippines
Quezon City, Philippines
celso.co@tip.edu.ph

Abstract—The processs variability covers typically the material, equipment, method, and man, the so-called 4 Ms'. A simple process of cutting a 10-meter rod into two equal halves is illustrated where all the 4Ms were treated. The simulation of variability for the 4 Ms' were discussed. The theoretical perspective has caveat on matter of mathematical expressions that are realistics. The algorith used was Monte Carlo Simulation based on random number generation that fit Gaussian Curve. The python programming language was used with symbolic math library, the sympy. Although the industry is used to 99.99%, for the sake tutotial the academe toys around 50%.

Index Terms—variability, Normal Distribution, Gausian, Man, Machine, Method, Material, 4Ms.

1 Introduction

The Monte Carlo Simulation requires the generation of random variable of which statistical distribution follows a certain distribution. In this case, the normal or Gaussian distribution is used based on the assertion that a stable process must obey central tendency theorem.

The Gaussian distribution function [4] is stated as follows.

$$f(x) = \frac{\sqrt{2}e^{-\frac{(\mu-x)^2}{2\sigma^2}}}{2\sqrt{\pi}\sigma} \tag{1}$$

where :
μ : Real number or a list representing the mean or the mean vector
σ : Real number or a positive definite square matrix,
$\sigma^2 > 0$ the variance
Returns: A Random Symbol.

The sympy function

$$f_1 = NormalDistribution(\mu, \sigma) \tag{2}$$

- *Technological Institute of the Philippines*

Manuscript received March 18, 2019

declares f_1 as the normal distribution function. The random variable of z of it is expressed as follows.

$$f_1(z) = \frac{\sqrt{2}e^{-\frac{(\mu-z)^2}{2\sigma^2}}}{2\sqrt{\pi}\sigma} \tag{3}$$

Integratinng (3), we have the following.

$$f_2(z) = \int_{z_i}^{z_f} f_1(z)\,dz = \frac{\operatorname{erf}\left(\frac{\sqrt{2}(-\mu+z_f)}{2\sigma}\right)}{2} - \frac{\operatorname{erf}\left(\frac{\sqrt{2}(-\mu+z_i)}{2\sigma}\right)}{2} \tag{4}$$

Let $\mu = 0$ and $\sigma = 1$ then f_2 from $z_i = -\infty$ to $z_f = \infty$ we have normal probability distribution as follows.

$$f_2(z) = \int_{-\infty}^{\infty} f_1(z)\,dz = 1 \tag{5}$$

The $f_3(z)$ from $z_i = -1$ to $z_f = 1$ implies lower limit at $-\sigma$ and upper limit σ.

$$\int_{-1}^{1} f_1(z)\,dz = \operatorname{erf}\left(\frac{\sqrt{2}}{2}\right) = 0.682689492137086 \tag{6}$$

At $\pm 3\sigma$,

$$\int_{-3}^{3} f_1(z)\,dz = \operatorname{erf}\left(\frac{3\sqrt{2}}{2}\right) = 0.99730020393674 \tag{7}$$

Finally at $\pm 6\sigma$,

$$\int_{-6}^{6} f_1(z)\,dz = \operatorname{erf}\left(3\sqrt{2}\right) = 0.999999998026825 \tag{8}$$

The defect per million DPM [1] of (6), (7), and (8) are given as follows.

$$DPM = 317310.507862914 \tag{9}$$

$$DPM = 2699.79606326021 \tag{10}$$

$$DPM = 0.00197317528982666 \tag{11}$$

2 Review of Related Literature

The context of this tutorial on process variability is the statistical process control (SPC) [2]. It is asserted that the stable process follows the central limit theorem [5]. Hence, by making the upper control limit and lower control limit of approaches a distant of 6σ from the mean μ, the process is made extremely stable nd predictable. The discussion would focus on the relationship of variables in processes to account for relative impact of assignable causes [3].

3 Simulation

Given a case where a 10m rod is to be cut into two halves where the statistiscs of random variables are those of machine, man, material and method, it is desired to find out how these affect the yield of the process. The objective of the analysis to improve the accuracy that is the distance of average from the desired mean and the precision that is standard deviation or variance.

3.1 Material

A set of values of z can be generated randomly by the sympy function
$Normal(\mu, \sigma)$. Given $\mu = 10$ and $\sigma = 1$ then for 10000 pieces we have

$$Rods = random.normal(nrodmu, nrodsigma, count) = [8.866765, 8.808836 \dots 8.053214]. \tag{12}$$

The histogram of (15) is shown in Figure 1.

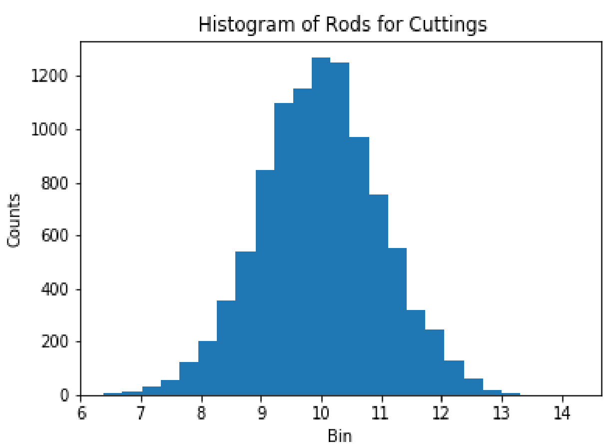

Fig. 1. 1. Histogram Rods Variation

Question:
Given the specification of ± 1, how many are good and what is the yield?
Answer number of good rods = 6954, yield = 69.54%. Note that everytime the program is run these values changes. The solution was an algorithm to check dimension of each rod and count those that are within tolerance. Solving it using (4) gives a unique answer unlike the results from the algorithm above. Why?

3.2 Equipment

Now consider that these rods are to be cut into halves. The cutter has $\mu = 0.5$ and $\sigma = 0.1$ then for 6954 pieces we have

$$f_1 = NormalDistribution(0.5, 0.1) \tag{13}$$

$$f_1(z) = \frac{5.0\sqrt{2}e^{-50.0(z-0.5)^2}}{\sqrt{\pi}} \tag{14}$$

$$f_2(z) = \int_{z_i}^{z_f} f_1(z)\, dz \tag{15}$$

$$Cut_{event} = random.normal(0.5, 0.1, 6954) = [0.5104791588224948, 0.5032845297161098, \dots 0.4233040541115318] \tag{16}$$

The histogram of (19) is shown in Figure 2.

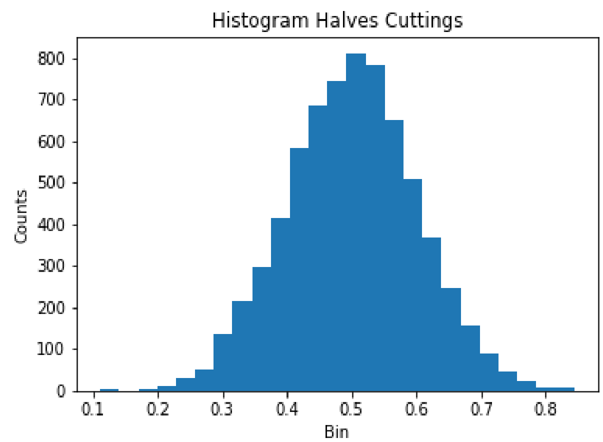

Fig. 2. Histogram of Cutter Variation

3.3 Method

There two methods to cut the rod. One method is the cutter is place at a distance of half the rod size exactly 5. Then the rod is referenced at the point of origin and cut is made. The other method is to place the rod at its center for cutter engagement. The formulas for the two methods are different.

Question: Determine the formula for each and justify it.
Answers:
1. The cutter with variation, i.e., $0.5 \pm .1$ is multiplied by the perfect dimension of rod that is 10. Why perfect 10? Then the answer is substracted from a rod with variation. The two answers are the cut products.
2. The rod with variation, i.e., 10 ± 1 is multiplied by the cutter with variation, i.e., $0.5 \pm .1$. The answer is substracted from the rod with variation. The two cuts are products.
The histograms of Method 1 and 2 are shown in Figure 3.

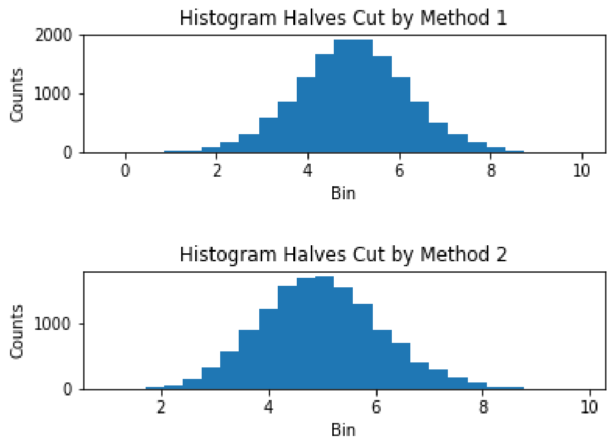

Fig. 3. Histogram of Cut Rods

Question:
Assuming that the cut specification is $5 \pm .5$, determine the number of good cut and the yield for each method
Answer:
1. The number of good cut for method 1 is 4557 while for method 2 4780 respectively.
2. The yield for method 1 is 32.7653% while for method 2 is 34.3687% respectively. Why is it method 2 has higher yield?

3.4 Operator

Assume and operator makes the loading with error of $\mu = 0.4$ and $\sigma = 0.1$. The histograms with operator handling for Method 1 and 2 are shown in Figure 4.

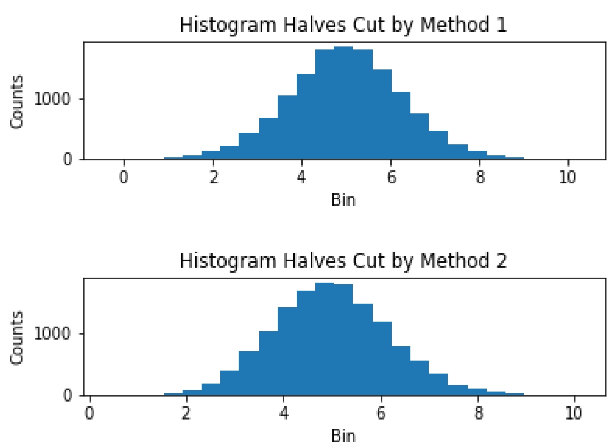

Fig. 4. Histogram of Cut Rods with Operator's Handling

The yield with operator's handling errors for method 1 is 31% while for method 2 is 32% respectively.

4 Discussion

4.1 Tolerance Variation

The yield numbers were exagerated for theoretical discussion. The four variables are material, equipment or machine, method, and operator or man. Hences, such sources of variations are known as 4M.

In real world, the overall yied is the only one observed and reported. It problematict to account the yield loss for each source. However, the so-called fool proofing refers to elimination of operator's mishandling. Tyically the solution is jig and fixture. In this case, the positional fixture with alignment jig could eliminate the operator's error.

The variation in machine could increase overtime. Hence, regular caligration is necessary. To improve the variability, the continuous improvement probram is the typical solution.

The material variation can be mitigated at a cost by simple requirement f0r tighter tolerances. However, the most difficult variation to estimate is the method. The difficulty was chosing the appropriate equation to define the variability in a method. For instance, either of the two methods has rejects of oversize. It is possible to introduct rework or reprocess to cut the oversize for yield recovery purposes. What are the appropriate equations and the corresponding algorithm that can be concocted for each method? If the target reference remains the mean, there is a chance that the outcome is undersize. Should it the reference be at maximum tolerance? In other words, the method for rework may not be the same as the standard process.

4.2 Other Type of Variations

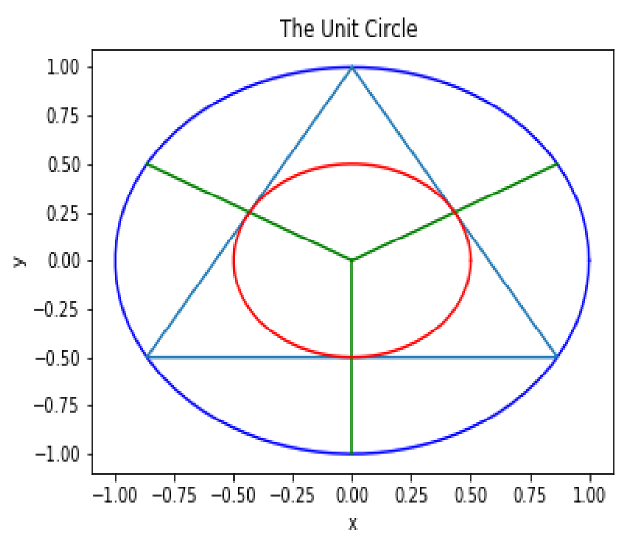

Fig. 5. Equilateral Triangle Inscribe in Unit Circle

Consider a unit circle where an equilateral triangle is inscribed as shown in Figure 5. Suppose a random line is to be drawn across it, determine the probability that the cord is less than the length of the the side of equilateral triangle. There are a number of answer depending on how the random lines are drawn.
1. The line drawn from any two points at the circumference,

gives the probability of $\frac{1}{3}$.
2. The line drawn perpendicular to radius from a any point along radius gives the probability of $\frac{1}{2}$
3. The line drawn tangent to any concentric circle gives the probabiity $\frac{1}{4}$.
4. From the length of the cord i.e. 0 to 2, the probability is $\frac{\sqrt{3}}{2}$.

The question is that which of the four probability numbers make sense?

5 Conclusion

The variation of process variables can simulated by pseudo random generator. The formula or equation associated with the algorithm must be ensured realistically. From theoretical perspective the variance of every variable can be determined but in real world operation, the net variance is the only one that can be measured. Hence, the defect modes may be difficult to itemize.

As discussed, the application of random variable depends on certain perspective of variation. For instance, a short cord drawn near the center of the circuit is not a line crossing the circle although it is random. It hsd to be moved from the center to have its end points touch the circumference.

Finally, the accuracy must be the difference between the average and the specified mean must be approaching zero. The precission or repeatability must be such that the tolerance is equal to a number of sixmas.

6 Recommendation

The Monte Carlo Simulation is useful for prediction of process yield. Due care should be excercised in formulation of equation and algorthm. The mathematics must be articulate the real world. There exists ample room for creativity in defect modes itemization. The design of experiments were used to optimize parameters. The Monte Carlo Simulation may help to reduce number of experiments.

References

[1] Sig Sigma Daily. Dpmo to sigma level relationship. *CopyrightÂ© 2019 Bisk. All rights reserved*, Mar 2019.
[2] International Datayzer. What is spc. *2015 Â© DATALYZER, All rights reserved*, Mar 2019.
[3] P. M. Swamidass, editor. *Assignable causes, of variability ASSIGNABLE CAUSES OF VARIATIONS*, pages 50–50. Springer US, Boston, MA, 2000.
[4] Sympy Development Team. Sympy documentation. *https://github.com/sympy/sympy/releases*, page 1117, Jul 2017. https://docs.sympy.org/latest/modules/stats.html#sympy.stats.Normal.
[5] Statistics How To. Central limit theorem. *Copyright Â©2019 Statistics How To | Theme by: Theme Horse | Powered by: WordPress*, Mar 2019.

About the Authors

Celso Bation Co He earned his diploma as Electronic Technician from International Correspondence Schools of University of Pennsylvania in 1972. He obtained his degree of Bachelor of Science in Electronics and Communication Engineering (ECE) from the University of Sto Tomas in 1977, Master and Doctoral degrees in ECE from De La Salle University in 1996 and 2007 respectively.
He is currently the research director of Technological Institute of the Philippines and an advocate of strong linkages among academes, industries and government institutions.